REMARKS

In an office action dated July 9, 2004, the Examiner rejected claims 1-3, 5, 7-9, 11, 13 and 16 under 35 U.S.C. 102(b) as anticipated by Brais et al. (U.S. Patent 5,995,936); rejected claims 4, 6, 10, 12 and 17 under 35 U.S.C. 103(a) as being unpatentable over *Brais* in view of Williams (U.S. Patent 6,308,154); and rejected claims 14-15 and 18-19 under 35 U.S.C. 103(a) as being unpatentable over *Brais* in view of Englehardt (U.S. Patent 5,477,511).

Applicant has cancelled claims 16-19, and the rejections thereof are moot. Applicant has amended the remaining independent claims 1, 5, 9 and 11 to more specifically recite the invention claimed herein. In particular, the claims have been amended to clarify the nature of the association of images with symbolic text which is performed automatically by the camera. Various dependent claims have been amended to conform to changed language in the independent claims. As amended, the claims are patentable over the cited art.

Applicant's invention relates to the use of digital cameras, and significantly, it relates to the ***user interface*** provided by a digital camera. As is well known in the field of user interface, a new and improved interface is not necessarily intended to provide any hitherto unavailable function, but to make available existing functions in a manner which is easier to learn, easier to remember, easier to manipulate, or in some other manner easier to use for some class of users.

The purpose of a digital camera is to capture digital images. Almost all commercially available digital cameras provide some means for uploading digital images to another device, such as a general purpose digital computer. Using a digital computer, any number of software applications provide the capability to generate captions and similar text associated with digital images, format the images and associated text, display or print the images and text, and so forth. Thus, it has always been possible to associate text (such as a caption) with a digital image.

Digital cameras were originally very expensive, and used only by professionals or very serious photographers. While existing tools are useful and applicant does not propose to dispense with them entirely, they are generally designed around the needs of these professionals or serious photographers, and are not ideally suited to the casual use of many users. Today, digital cameras are a commodity item which are rapidly replacing standard film cameras. Many if not most of the users of digital cameras today know relatively little about the science of photography, and simply want a camera which they can point at a subject and take a picture. Many of these users also know relatively little about digital computers, or the image editing or report generating software available on digital computers.

Manufacturers of digital cameras have recognized this market, and have designed many of their camera models around such users. For example, many digital cameras offered for sale today include an automated mode wherein the camera automatically selects a focal length, aperture, exposure time, and other parameters. Many such cameras also include various scene shooting modes, such as a portrait mode, a landscape mode, and so forth. In such cameras, simplicity is the key to the interface. Frequently, such cameras simplify the interface by not allowing the user to directly set the camera's parameters, but only to select one of the automatic operating modes of the camera. With such an interface, an amateur photographer can take a reasonably good photograph knowing almost nothing about the science of photography.

Applicant's invention is in the spirit of such an automated interface. The casual user may frequently wish to record separate information associated with each of multiple digital images for later reference. Using existing tools, it is possible to separately record information and to subsequently integrate it with the images using various tools available on a digital computer. But this is too cumbersome for the average user. Applicant's invention provides this capability in a very simple interface. In accordance with applicant's preferred embodiment, the user simply pushes a button to record information by speaking into a microphone in the camera, and the

camera ***automatically makes an association*** between the recorded information and one of the images. This association is recorded in the memory of the camera, so that the image and the associated text are automatically paired with each other. Ideally, software to print or display images would offer a function to print or display the associated text with the image automatically, without requiring the user to perform any additional editing function. It would alternatively be possible to edit either the image or associated text using conventional editing tools.

The original independent claims were insufficiently specific regarding the essential feature of automatically making an association between the image and the text. Applicant has accordingly amended all independent claims to clarify this feature. As amended, the claims recite that the user speaks discrete segments of human speech, and the camera automatically associates each discrete segment with a respective digital image based on the time the speech segment is spoken. E.g., the camera automatically associates the speech with a digital image taken at approximately the same time. Representative amended claim 1 recites in part:

1. A digital camera, comprising:
a housing;
a digital optical sensing apparatus ...;
a storage medium for storing digital optical images ...;
an acoustic sensor capable of sensing human speech;
a speech reduction apparatus ... converting human speech ... to a symbolic text form; and
a controller which *stores said symbolic text form in said storage medium in a relationship associated with a captured digital image*, wherein said controller:
(a) receives a user indication of a plurality of discrete time intervals;
(b) records a plurality of discrete human speech segments sensed by said acoustic sensor in respective said discrete time intervals;
(c) causes said speech reduction apparatus to convert each said human speech segment to a corresponding symbolic text segment; and
(d) *automatically associates a respective digital optical image captured by said digital optical sensing apparatus with each said symbolic text segment based on a temporal relationship between the time interval in which the discrete human speech segment corresponding to the symbolic text segment was recorded and the capturing of said digital optical image.* [emphasis added]

The remaining independent claims, although not identical in scope, contain limitations analogous to the italicized language above.

Brais, cited by the Examiner, discloses a report generation system, in which a digital camera, microphone and portable computer are linked and attached to a user, and the user records both digital images and text while observing something of interest, such as an industrial facility being inspected. In accordance with *Brais*, the system "associates" recorded images and voice commentary, using several possible techniques. In one technique, the image is associated (via a timestamp) with a specific location in the audio recording corresponding to the time at which the image is taken. An association can also be made by means of specific editing commands from the user.

Although *Brais*' "association" of images and text may appear to be similar to applicants, a careful consideration of the claims as amended shows that they are quite different. *Brais*' system essentially puts conventional editing software in a portable computing device, providing conventional editing functions to the user at the same time that the user is capturing images. While it is possible to create an association between text and an image using such conventional editing functions (e.g., defining an image caption) and to do so at approximately the same time that the image is captured, such functions inherently require user input. I.e., these editing function do not "automatically associate" an image with text based on a temporal relationship between the time the image is captured and the time the speech segment is recorded, as required by applicant's claims.

Brais does provide one automatic association function, and that is to associate an image with a specific location in an audio recording. However, in this case the image is not associated with any particular text, but with a location in the text. Applicant's claims recite the association of a discrete segment of text with an image, and this is a subtle but important difference.

Applicant will address the issue of obviousness presently, but since the Examiner's rejection was for anticipation, applicant wishes to make this point very clear. *Brais* automatically associates an image with a ***location*** within a body of text. Multiple images may be associated with different respective locations within the same body of text. Applicant's claims recite that each of multiple images is associated with a ***discrete segment*** of a plurality of discrete segments of text. Although any arbitrary discrete segment of *Brais*' text might be closer to one image than another, the system does not define any discrete segments, nor does it associate them with corresponding images. To the extent such an association is made, it must be made manually by the user. Accordingly, the claims as amended are not anticipated by *Brais*.

Nor are the claims obvious over *Brais*. Even the most casual reader will observe that *Brais*' system is fundamentally different from that claimed by applicant. *Brais*' system is intended to be a report generating system, and the primary purpose of the system is the collection of text. The digital images are merely ancillary to the text, and for this reason are associated with locations within the text. This makes perfect sense in the context of what *Brais* is attempting to accomplish. There would be no motivation to apply applicant's system to the report generating environment of *Brais*. An inspector might well dictate a large number of comments before deciding to photograph something, and by automatically associating the photograph which all of the dictated text, the context of the photograph could well be lost.

Far from suggesting applicant's invention, *Brais* teaches away from it. *Brais* recognizes that in some circumstances it may be desirable to associate a caption, i.e. a discrete segment of text, with an image. How does *Brais* do so? By use of conventional editing functions, whereby a user explicitly tells the system to associate certain text with the image. *Brais* recognizes that in their environment, it would make little sense to be making automatic associations of discrete text segments to images, and therefore reserves this capability to the user.

Applicant's invention is intended to improve the user interface for the casual user of a digital camera. In this environment, the user is primarily generating digital images; to the extent text is generated, it is only ancillary to this primary function. The casual user does not want to be bothered with complicated function keys and combinations for associating text with images. Experience has shown that most casual users simply will not use this capability, and the information will go unrecorded. In order to support the casual user, the capability to associate captions (discrete segments of text) with images should be automatic. In accordance with applicant's invention, the user dictates a discrete segment and the camera automatically associates it with an image. What could be simpler?

Williams, a secondary reference, is cited to show the encoding of human speech as phonemes according to certain dependent claims herein, and does not teach or suggest any particular interface in a digital camera. *Englehardt*, another secondary reference, is cited to show printing and viewing of digitally generated images and text. *Englehardt* discloses a portable dictation system similar in certain ways to that of *Brais*. *Englehardt* discloses that a digital image array can be coupled to the dictation system, but does not teach or suggest any particular interface in a digital camera, and in particular does not teach or suggest automatically associating an image with a discrete segment of text as claimed by applicant.

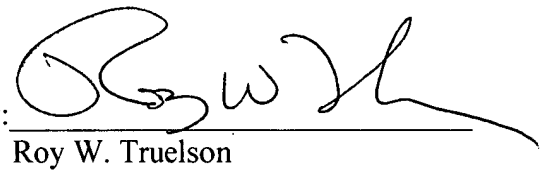
In hindsight, it is easy to say that the interface claimed by applicants is not rocket science, and that anyone might have thought of such a feature. But hindsight is not the proper test. In the realm of user interface, where simplicity is often the ultimate invention, many inventions which simplify the interface look easy in hindsight. But nothing in the cited art, and certainly not *Brais*, suggests the association of discrete text segments with corresponding images performed automatically by the camera, as disclosed and claimed by applicants. For all of the above reasons, the claims as amended are patentable over the cited art.

Applicant has added several dependent claims which recite more particularly the priority scheme used by the camera to associate discrete text segments with digital images. This priority scheme is disclosed beginning at p. 10, line 8 of the specification. No new matter is introduced.

In view of the foregoing, applicant submits that the claims are now in condition for allowance and respectfully requests reconsideration and allowance of all claims. In addition, the Examiner is encouraged to contact applicant's attorney by telephone if there are outstanding issues left to be resolved to place this case in condition for allowance.

Respectfully submitted,

PAUL S. HALVORSON

By: 

Roy W. Truelson
Registration No. 34,265

Telephone: (507) 289-6256